

# Patent Abstracts of Japan

PUBLICATION NUMBER : 09233231

PUBLICATION DATE : 05-09-97

APPLICATION DATE : 20-02-96

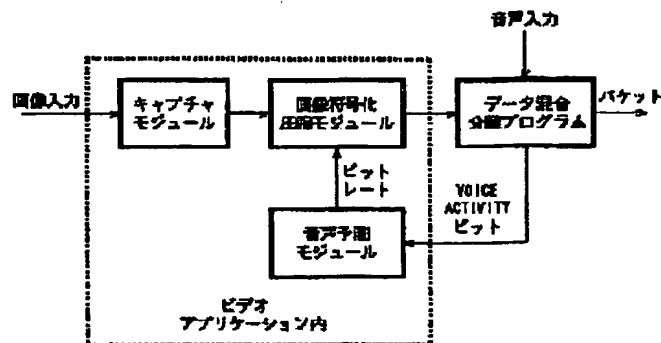
APPLICATION NUMBER : 08031578

APPLICANT : INTERNATL BUSINESS MACH CORP  
<IBM>;

INVENTOR : ISHIKAWA KATSUYA;

INT.CL. : H04M 11/06 H04L 12/56 H04N 7/24  
H04N 7/15

TITLE : DATA TRANSMISSION METHOD AND  
DEVICE THEREFOR



ABSTRACT : PROBLEM TO BE SOLVED: To send also image data without residence regardless of placing priority onto voice data by adjusting a compression rate of the image data depending on whether or not a packet includes the voice data.

SOLUTION: When image data are received, an image code compression rate module applies code compression to the data and a data mixing separate program takes over the control. The data mixing separate program applies code compression to the voice data received at this point of time and mixes the voice data with the received image data and divides the resulting data into packets and updates a Voice Activity (VA) bit denoting the presence of voice input for each 20msec. A voice prediction module reads the VA bit for each prescribed time (t) and references a history of the VA bit within a prescribed time T (T>t) to predict whether or not the voice input is received while image data of a frame are transferred, and also calculates an optimum bit rate based on the result of prediction and feeds back the result.

COPYRIGHT: (C)1997,JPO

(19) 日本国特許庁 (J P)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平9-233231

(43) 公開日 平成9年(1997)9月5日

(51) Int.Cl. <sup>6</sup>	識別記号	庁内整理番号	F I	技術表示箇所
H 0 4 M 11/06			H 0 4 M 11/06	
H 0 4 L 12/56			H 0 4 N 7/15	
H 0 4 N 7/24		9466-5K	H 0 4 L 11/20	1 0 2 A
7/15			H 0 4 N 7/13	Z

審査請求 未請求 請求項の数18 O L (全 17 頁)

(21) 出願番号 特願平8-31578

(22) 出願日 平成8年(1996)2月20日

(71) 出願人 390009531

インターナショナル・ビジネス・マシーンズ・コーポレーション

INTERNATIONAL BUSINESS MACHINES CORPORATION

アメリカ合衆国10504、ニューヨーク州

アーモンク (番地なし)

(72) 発明者 石 川 克 也

神奈川県大和市下鶴間1623番地14 日本ア

イ・ビー・エム株式会社 大和事業所内

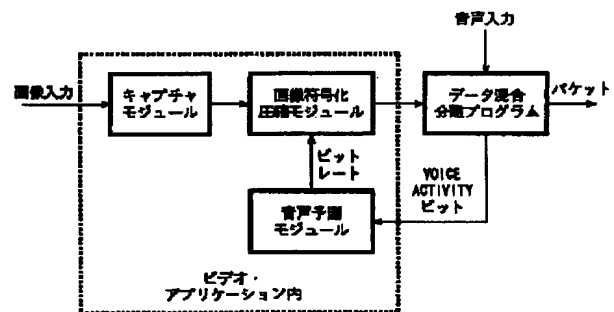
(74) 代理人 弁理士 合田 潔 (外2名)

## (54) 【発明の名称】 データ伝送方法及び装置

## (57) 【要約】 (修正有)

【課題】 画像データに割り当てられる帯域幅が変動しても、好適に伝送できるようにする。

【解決手段】 符号化圧縮した音声データと画像データを固定長のバケットに混合して伝送するタイプのデータ伝送方法において、(a) 音声データの投入を試みる段階と、(b) 音声データの投入の有無を検出する段階と、(c) 画像データを所定の捕捉間隔で捕捉する段階と、(d) 画像データを所定の圧縮率で符号化圧縮する段階と、(e) 前記(b)段階の検出結果に応じて、音声データを符号化圧縮し、符号化圧縮された画像データと混合してバケットを生成する段階と、(f) 生成したバケットを伝送する段階と、(g) 過去1回以上の前記(b)段階における検出結果を基に、以後の音声データの有無を予測する音声予測段階と、(h) 前記(g)段階の予測結果を基に、前記(c)段階の前記所定の捕捉間隔を調整する段階と、を具備する。



**【特許請求の範囲】**

【請求項1】符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送方法において、パケット内に音声データを含むかどうかに応じて画像データの圧縮率を調整する、ことを特徴とするデータ伝送方法。

【請求項2】入力した音声データを符号化圧縮するとともに、捕捉した画像データを符号化圧縮し、符号化圧縮された音声データと符号化圧縮された画像データを固定ビット幅のパケットに混合して、ネットワーク上に伝送するタイプのデータ伝送方法において、パケット内に音声データを含むかどうかに応じて画像データの捕捉間隔を調整する、ことを特徴とするデータ伝送方法。

【請求項3】符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送方法において、(a) 音声データの入力を試みる段階と、(b) 音声データの入力の有無を検出する段階と、(c) 画像データを所定の捕捉間隔で捕捉する段階と、(d) 画像データを所定の圧縮率で符号化圧縮する段階と、(e) 前記(b)段階の検出結果に応じて、音声データを符号化圧縮するとともに、符号化圧縮された画像データと混合して、パケットを生成する段階と、(f) 生成したパケットを伝送する段階と、(g) 過去1回以上の前記(b)段階における検出結果を基に、以後の音声データの有無を予測する音声予測段階と、(h) 前記(g)段階の予測結果を基に、前記(c)段階の前記所定の捕捉間隔を調整する段階と、を具備することを特徴とするデータ伝送方法。

【請求項4】前記音声予測段階では、過去の所定期間内で複数回音声入力の有無を調べ、音声入力が所定の割合以上あれば将来音声入力があると判断し、所定の割合未満であれば将来音声入力がないと判断する、ことを特徴とする請求項3に記載のデータ伝送方法。

【請求項5】前記音声予測段階では、過去の所定期間内で複数回調べた音声入力の有無を時間的に加重平均し、加重平均値が閾値以上であれば将来音声入力があると判断し、閾値未満であれば将来音声入力がないと判断する、ことを特徴とする請求項3に記載のデータ伝送方法。

【請求項6】前記(c)段階では比較的短い第1の捕捉間隔又は比較的長い第2の捕捉間隔のうちのいずれか一方で画像データを捕捉し、前記(h)段階では、予測結果が音声データ有りを示す場合は第2の捕捉間隔を選択し、予測結果が音声データ無しを示す場合は第1の捕捉間隔を選択する、ことを特徴とする請求項3に記載のデータ伝送方法。

【請求項7】符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネッ

トワーク上に伝送するタイプのデータ伝送方法において、(a) 音声データの入力を試みる段階と、(b) 音声データの入力の有無を検出する段階と、(c) 画像データを所定の捕捉間隔で捕捉する段階と、(d) 画像データを所定の圧縮率で符号化圧縮する段階と、(e) 前記(b)段階の検出結果に応じて、音声データを符号化圧縮するとともに、符号化圧縮された画像データと混合して、パケットを生成する段階と、(f) 生成したパケットを伝送する段階と、(g) 過去1回以上の前記(b)段階における検出結果を基に、以後の音声データの有無を予測する音声予測段階と、(h) 前記(g)段階の予測結果を基に、前記(d)段階の前記所定の圧縮率を調整する段階と、を具備することを特徴とするデータ伝送方法。

【請求項8】前記(d)段階では、比較的高い第1の圧縮率又は比較的低い第2の圧縮率のうちのいずれか一方で画像データを圧縮し、前記(h)段階では、予測結果が音声データ有りを示す場合は第1の圧縮率を選択し、予測結果が音声データ無しを示す場合は第2の圧縮率を選択する、ことを特徴とする請求項7に記載のデータ伝送方法。

【請求項9】前記音声予測段階では、過去の所定期間内で複数回音声入力の有無を調べ、音声入力が所定の割合以上あれば将来音声入力があると判断し、所定の割合未満であれば将来音声入力がないと判断する、ことを特徴とする請求項7に記載のデータ伝送方法。

【請求項10】前記音声予測段階では、過去の所定期間内で複数回調べた音声入力の有無を時間的に加重平均し、加重平均値が閾値以上であれば将来音声入力があると判断し、閾値未満であれば将来音声入力がないと判断する、ことを特徴とする請求項7に記載のデータ伝送方法。

【請求項11】符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送装置において、(a) 音声データの入力を試みる音声入力手段と、(b) 音声データの入力の有無を検出する音声検出手段と、(c) 画像データを入力する画像入力手段と、(d) 入力した画像データを所定の捕捉間隔で捕捉する画像捕捉手段と、(e) 画像データを所定の圧縮率で符号化圧縮する画像符号化圧縮手段と、(f) 前記音声検出手段による検出結果に応じて、音声データを符号化圧縮するとともに、符号化圧縮された画像データと混合して、パケットを生成するデータ混合手段と、(g) 生成したパケットを伝送する伝送手段と、(h) 前記音声検出手段による過去1回以上の検出結果を基に、以後の音声データの有無を予測する音声予測手段と、(i) 音声予測手段による予測結果に応じて、前記画像捕捉手段の前記所定の捕捉間隔を調整する調整手段と、を具備することを特徴とするデータ伝送装置。

【請求項12】前記画像捕捉手段は比較的短い第1の捕捉間隔又は比較的長い第2の捕捉間隔のうちのいずれか一方で画像データを捕捉し、前記調整手段は、予測結果が音声データ有りを示す場合は第2の捕捉間隔を選択し、予測結果が音声データ無しを示す場合は第1の捕捉間隔を選択する、ことを特徴とする請求項11に記載のデータ伝送装置。

【請求項13】前記音声予測手段は、過去の所定期間内で複数回音声入力の有無を調べ、音声入力が所定の割合以上あれば将来音声入力があると判断し、所定の割合未満であれば将来音声入力がないと判断する、ことを特徴とする請求項11に記載のデータ伝送装置。

【請求項14】前記音声予測手段は、過去の所定期間内で複数回調べた音声入力の有無を時間的に加重平均し、加重平均値が閾値以上であれば将来音声入力があると判断し、閾値未満であれば将来音声入力がないと判断する、ことを特徴とする請求項11に記載のデータ伝送装置。

【請求項15】符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送装置において、(a) 音声データの入力を試みる音声入力手段と、(b) 音声データの入力の有無を検出する音声検出手段と、(c) 画像データを入力する画像入力手段と、(d) 入力した画像データを所定の捕捉間隔で捕捉する画像捕捉手段と、(e) 画像データを所定の圧縮率で符号化圧縮する画像符号化圧縮手段と、(f) 前記音声検出手段による検出結果に応じて、音声データを符号化圧縮するとともに、符号化圧縮された画像データと混合して、パケットを生成するデータ混合手段と、(g) 生成したパケットを伝送する伝送手段と、(h) 前記音声検出手段による過去1回以上の検出結果を基に、以後の音声データの有無を予測する音声予測手段と、(i) 音声予測手段による予測結果に応じて、前記画像符号化圧縮手段の前記所定の圧縮率を調整する調整手段と、を具備することを特徴とするデータ伝送装置。

【請求項16】前記画像符号化圧縮手段は比較的高い第1の圧縮率又は比較的低い第2の圧縮率のうちのいずれか一方で画像データを捕捉し、前記調整手段は、予測結果が音声データ有りを示す場合は第1の圧縮率を選択し、予測結果が音声データ無しを示す場合は第2の圧縮率を選択する、ことを特徴とする請求項15に記載のデータ伝送装置。

【請求項17】前記音声予測手段は、過去の所定期間内で複数回音声入力の有無を調べ、音声入力が所定の割合以上あれば将来音声入力があると判断し、所定の割合未満であれば将来音声入力がないと判断する、ことを特徴とする請求項15に記載のデータ伝送装置。

【請求項18】前記音声予測手段は、過去の所定期間内で複数回調べた音声入力の有無を時間的に加重平均し、

加重平均値が閾値以上であれば将来音声入力があると判断し、閾値未満であれば将来音声入力がないと判断する、ことを特徴とする請求項15に記載のデータ伝送装置。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、パケット形式でデータを伝送するタイプのデータ伝送方法及び装置に係り、特に、ビデオ会議のための音声データ及び画像データをパケット形式にして伝送するためのデータ伝送方法及び装置に関する。更に詳しくは、本発明は、音声データ及び画像データを多重化したパケット形式にして伝送するためのデータ伝送方法及び装置に関する。

【0002】

【従来の技術】

ビデオ会議システム：最近、パーソナル・コンピュータ（PC）が広範に普及し、事務用及び家庭用のユーザ間で相当に一般的なものとなってきている。これに伴って、その用途も拡大し、単なるOA機器としてだけではなく、情報交換の媒体としても用いられるようになってきた。例えば、遠隔にある会議室間を通信回線によって結び、授受される音声データや画像データをPCで処理するという、いわゆるビデオ会議システム（若しくはテレビ会議システム）が注目を集めるようになってきた。

【0003】このようなビデオ会議システムには、通信媒体としてISDN（Integrated Services Digital Network）を、処理装置としてデスクトップPCを用いるのが主流であった。ISDNは、1本の通信線に対して音声伝送用及びデータ伝送用の2種類のチャネルを論理的に割り当てることができる、デジタル・データ伝送用のネットワークである。すなわち、ISDNは、電話などの音声に限らず、テキスト、データ、静止画、動画などのいわゆるマルチメディアを総合的に扱うことができる伝送媒体なのである。また、デスクトップPCが主流なのは、普及率が高いことその他、ビデオ会議の各参加者がそれぞれのオフィスの定められた場所に居ることが前提であったことにも依拠する。

【0004】ところが、昨今の目覚ましい技術革新に伴って、小型・軽量の、いわゆるノートブックPCが出現してきた。殆ど全てのノートブックPCは、バッテリー駆動型であり、屋外での携帯的・可搬的な使用、すなわちモバイル環境での使用が可能となっている。これに伴って、モバイル環境下でのビデオ会議に対するニーズも高まりつつある。

【0005】モバイル環境下でビデオ会議を実現する場合、データ伝送媒体として何を用いるか、ということが1つの問題となる。先述したISDNは、高機能な反面、高価で普及率も未だ高くない。したがって、ISDNによれば接続個所がかなり限定されてしまい、モバイル性を損なうことになる。これに対して、一般公衆回

線(POTS: Plain Old Telephone System)は、安価で普及率が高い。会議の各参加者は、自身のノートブックPCを、モデムを介して最寄の場所から一般公衆回線に接続することができる。したがって、ビデオ会議の通信媒体として、一般公衆回線への期待が必然的に膨らむ訳である。

【0006】図9には、一般公衆回線とPCを用いたビデオ会議ネットワークの構成を模式的に示している。同図に示すように、各PCはモデムを介して一般公衆回線網に接続される。また、PBX(Private Branch eXchange: 構内交換機)を経由して構内回線網上にPCが接続されている場合もある。なお、図示していないが、各PCには、ユーザの姿態等を撮像するためのビデオ・カメラ、入力映像をデジタル化してコンピュータ内に取り込むためのビデオ・キャプチャ・ボード/コントローラ、音声を入出力するためのマイク及びスピーカ、音声入出力をデジタル処理するためのオーディオ・コントローラ等、ビデオ会議に必須のハードウェア構成要素も備えられている。

【0007】一般公衆回線を用いてビデオ会議を実現する場合、データ転送量が最大の問題となる。何故ならば、ビデオ会議のためには、音声や画像を含むため、伝送しなければならない総データ量は、1本の回線に許容される帯域幅(すなわち最大の転送レート)よりもはるかに大きいからである。一般公衆回線を用いたビデオ会議システムの過去の例では、(1)音声の伝送を諦める、あるいは、(2)音声データと画像に各1本ずつ専用の回線を使用する、などの安直なソリューションが採られていた。しかし、最近では、音声及び画像についてのデータ圧縮技術やCPUの処理能力が向上したことに伴って、1本の公衆回線だけで音声データと画像データの両方を混合(若しくは多重化)して、伝送することが可能になってきた。

【0008】パケット方式による音声データ・画像データの伝送: 回線上でのコミュニケーションは、一般には、パケット通信、すなわち一連のデータを固定ビット幅のパケット単位に分割して行われる。ここで、パケットとは、伝送データの実体を含んだデータ部と、伝送データの属性情報などからなるヘッダ部とで構成される。通常、音声データや画像データは、パケットに分割する前に、それぞれ符号化圧縮されている。

【0009】ビデオ会議のための音声データ及び画像データを一本化して伝送する場合、音声データの方にプライオリティを与えるべきと考えられる。何故ならば、音声の途切れは不快感を与えるのみならず会話自体を不能にするため、音声により高いリアルタイム性が要求されるからである。このため、音声データと画像データを同時に送る場合には、パケット中の帯域を先ず音声データに割り当て、残りを画像データに割り当てられる。但し、このことは、同じ通信路を共有する画像データに対

して遅延を強要することをも意味している。

【0010】図10には、音声データ及び画像データを伝送するためのパケット構造の一例を示している。1つのパケットは288ビット長である。これは、最大転送レートが14.4kbpsのモデムを用いた場合の、20(=50分の1)msec分のデータ量に相当する。パケット内のデータ・フィールドの割り振り方は、音声データを含むかどうかによって2種類に分かれる。

【0011】図10(a)には、音声データ付きパケット(「VOD(Voice Over Data)パケット」ともいう)の構造を示している。第1ビットは、SYNCであり、同期を採るために用いられる。第2ビットは、該パケット中に音声データを含んでいるかどうかを示すためのGSMビットであり、音声データを含む場合には音声フラグ(「Voice Activityビット(音声入出力監視ビット)」ともいう)が設定(ON)される。SYNCビットとGSMビットが、パケットのヘッダ部であり、第3ビット以降がデータ部である。第3ビットから第8ビットの6ビットは、CRC(Cyclic Redundancy Check)、すなわち、送信データのエラー検出のために用いられる。第9～第272の264ビットは、音声データに割り当てられる(但し、264ビット中の4ビットはパリティ・ビットとして使用される)。伝送される音声データは、例えばGSM/\*アルゴリズムによって符号化圧縮されている。残りの第273～第288ビットの16ビットは、画像データに割り当てられる。画像データは、例えばMPEG(Motion Picture Experts Group)1やH261\*\*によって符号化圧縮されている。このパケットによれば、最大13kbps(=260ビット÷20msec)の転送レートで音声データが伝送される。

【0012】また、図10(b)には、音声データを含まないパケット(「NON VOICEパケット」ともいう)の構造を示している。第1ビットは、SYNCであり、同期を採るために用いられる。第2ビットは、該パケット中に音声データを含んでいるかどうかを示すためのGSMビットであり、音声データを含まない場合にはVoice Activityフラグが解除(OFF)される。SYNCビットとGSMビットが、パケットのヘッダ部であり、第3ビット以降がデータ部である。この場合、残りの第3～第288ビットの286ビットの帯域全てが画像データに割り当てられる。画像データは、前述同様、MPEG1やH261に基づいて符号化圧縮されている。

【0013】パケット通信に伴う問題点: 音声データと画像データを多重化して伝送する場合、前述したように、リアルタイム性を要求される音声データの方にプライオリティを与えられる。このため、音声データがパケット中に存在するかどうかによって、画像データに割り当てられる帯域幅が大きく異なってくる。このことは図

10を参照すれば、直感的に理解できよう。パケット中の画像データの帯域幅が変動することの帰結として、以下のような問題点が派生する。

#### 【0014】(1) 画像データの符号化圧縮のビット・レートの問題

画像データの符号化圧縮モジュール(ソフトウェア)、若しくは動画圧縮器(ハードウェア)は、一般には、自身に与えられたパラメータ(例えばビット・レート)に従って、データ圧縮率を調整するようになっている。より具体的には、ビット・レートに応じて、単位時間당りに符号化圧縮処理する画像フレーム数を一定に保つように働く。したがって、適切なビット・レートを与えることによって、画像データの転送レート(帯域幅)と画質のバランスのとれた最適なデータ転送を行うことができる。しかしながら、上述のように画像データに割り当てられる帯域幅が動的に変化すると、これに応じて最適なビット・レートも変動してしまう。音声データがパケット中に存在しない(すなわち画像データに与えられた帯域幅が広い)ことを前提にして、予め大きなビット・レートを与えてしまうと、画質はよくなるが、画像1フレーム分のデータ量が膨大になってしまう。動画の圧縮/伸長モジュール、若しくは圧縮/伸長器は、一般には1フレーム単位で画像データを捌くようになっている。したがって、1フレーム分のデータ量が大きくなり、受信側で画像1フレームを受信するための所要時間が長くなれば、その分画像データの伸長や表示する時期も遅滞する。この結果、受信者は数秒前の画像を見せられることになる。逆に、音声データがパケット中に存在する(すなわち画像データに与えられた帯域幅が狭い)ことを前提にして、予め小さなビット・レートを与えれば、画像1フレーム分のデータ量が少なくなるので、画像の遅延は解消される。しかし、そのトレードオフとして、音声データがなく広い帯域を与えられている間も、画像は貧弱なものになる。

#### 【0015】(2) ビデオ・キャプチャのフレーム・レートの問題

PCをビデオ会議に利用するには、通常、ビデオ・キャプチャ・ボード若しくはビデオ・キャプチャ・コントローラなどのような、ビデオ・カメラからの入力画像をデジタル化してファイルに落とすためのデバイスが用いられる。ビデオ・キャプチャ・コントローラは、一般には、1フレーム単位で画像データのキャプチャリングを行う。キャプチャリング動作は、例えば、上位のハードウェア(例えば「ビデオ・アプリケーション・プログラム」を実行するCPU)からの画像入力要求に回答して行われる。画像の動きをなるべくスムーズに見せるためには、フレーム・レート、すなわち1秒당りにキャプチャする画像フレーム数を上げればよい。しかし、その当然の帰結として、画像データの総量が増えることになる。狭い帯域幅しか与えられていないとき(図10

(a)参照)には、フレーム・レートを高くしていても、データの滞り(バッファリング)が生じ、受信者には数秒前の画像を見せられることになる。逆に、フレーム・レートを下げ過ぎると、画像の遅延は避けられるが、見た目の動画のスムーズさを欠くことになる。また、次フレームとの間にビデオ・データが送られない空白のパケット(隙間)が生じるなど、効率的な伝送を行えない。1フレーム以上のデータ・バッファリングは、画像の遅延を将来するだけで全く意味がない。つまり、1フレームの伝送を完了したときに次の1フレームの伝送を開始できるような間隔でキャプチャリングするのが望ましい。画像1フレーム分のデータ量(ビット数)と通信路中で画像データに割り当てられた帯域幅とから1フレーム分の伝送時間を計算すれば、次に画像フレームをキャプチャすべき最適な時間間隔が求められよう。しかしながら、画像データに割り当てられる帯域幅が動的に変化したのでは、この計算は成り立たない。

【0016】音声データの方にプライオリティが与えられるといっても、画像も一定の速度で且つ極力滑らかに再生することが望ましい。したがって、画像データの符号化圧縮や、ビデオ・キャプチャリングの問題の解決は不可避なのである。

【0017】ISDNやLANのように、音声データと画像データの各々に対して広い帯域幅の通信路を割り当てることができるような伝送方式であれば、上述した問題点は目立たない。単一の電話回線のように狭帯域の1通信路をデータ・チャンネルが共有するような伝送方式の場合、このような問題点は、とりわけ重大となる。

#### 【0018】《注釈》

\* : GSMは、Global System for Mobile communicationの略である。GSMにおける音声符号化アルゴリズムは、Regular Pulse Excited-Linear Predictive Code r(RPE-LPC)として知られている。

\*\* : H261はITU(International Telecommunication Union)勧告に従った圧縮アルゴリズム

#### 【0019】

【発明が解決しようとする課題】本発明の目的は、ビデオ会議のためのデータをパケット形式で伝送するための、優れたデータ伝送方法及び装置を提供することにある。

【0020】本発明の更なる目的は、ビデオ会議の音声データ及び画像データを多重化したパケット形式にして伝送するための、優れたデータ伝送方法及び装置を提供することにある。

【0021】本発明の更なる目的は、符号化圧縮した音声データと符号化圧縮した画像データを多重化して固定ビット幅のパケットにして伝送するときに、音声データにプライオリティを与えながらも画像データも滞りなく伝送することができる、優れたデータ伝送方法及び装置を提供することにある。

【0022】本発明の更なる目的は、符号化圧縮した画像データを固定ビット幅のパケット内に入れて伝送する場合において、画像データに割り当てられる帯域幅が変動しても、好適に画像データを伝送することができる優れたデータ伝送方法及び装置を提供することにある。

【0023】

【課題を解決するための手段】本発明は、上記課題を参酌してなされたものであり、その第1の側面は、符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送方法において、パケット内に音声データを含むかどうかに応じて画像データの圧縮率を調整する、ことを特徴とするデータ伝送方法である。

【0024】また、本発明の第2の側面は、入力した音声データを符号化圧縮するとともに、捕捉した画像データを符号化圧縮し、符号化圧縮された音声データと符号化圧縮された画像データを固定ビット幅のパケットに混合して、ネットワーク上に伝送するタイプのデータ伝送方法において、パケット内に音声データを含むかどうかに応じて画像データの捕捉間隔を調整する、ことを特徴とするデータ伝送方法である。

【0025】また、本発明の第3の側面は、符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送方法において、(a) 音声データの入力を試みる段階と、(b) 音声データの入力の有無を検出する段階と、(c) 画像データを所定の捕捉間隔で捕捉する段階と、(d) 画像データを所定の圧縮率で符号化圧縮する段階と、(e) 前記(b)段階の検出結果に応じて、音声データを符号化圧縮するとともに、符号化圧縮された画像データと混合して、パケットを生成する段階と、(f) 生成したパケットを伝送する段階と、

(g) 過去1回以上の前記(b)段階における検出結果を基に、以後の音声データの有無を予測する音声予測段階と、(h) 前記(g)段階の予測結果を基に、前記(c)段階の前記所定の捕捉間隔を調整する段階と、を具備することを特徴とするデータ伝送方法である。

【0026】ここで、前記(c)段階では比較的短い第1の捕捉間隔又は比較的長い第2の捕捉間隔のうちのいずれか一方で画像データを捕捉し、また、前記(h)段階では、予測結果が音声データ有りを示す場合は第2の捕捉間隔を選択し、予測結果が音声データ無しを示す場合は第1の捕捉間隔を選択する、ようにしてもよい。

【0027】また、本発明の第4の側面は、符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送方法において、(a) 音声データの入力を試みる段階と、(b) 音声データの入力の有無を検出する段階と、(c) 画像データを所定の捕捉間隔で捕捉する段階と、(d) 画像データを所定の圧縮率で符号

化圧縮する段階と、(e) 前記(b)段階の検出結果に応じて、音声データを符号化圧縮するとともに、符号化圧縮された画像データと混合して、パケットを生成する段階と、(f) 生成したパケットを伝送する段階と、

(g) 過去1回以上の前記(b)段階における検出結果を基に、以後の音声データの有無を予測する音声予測段階と、(h) 前記(g)段階の予測結果を基に、前記(d)段階の前記所定の圧縮率を調整する段階と、を具備することを特徴とするデータ伝送方法である。

【0028】ここで、前記(d)段階では、比較的高い第1の圧縮率又は比較的低い第2の圧縮率のうちのいずれか一方で画像データを圧縮し、また、前記(h)段階では、予測結果が音声データ有りを示す場合は第1の圧縮率を選択し、予測結果が音声データ無しを示す場合は第2の圧縮率を選択する、ようにしてもよい。

【0029】また、本発明の第5の側面は、符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送装置において、(a) 音声データの入力を試みる音声入力手段と、(b) 音声データの入力の有無を検出する音声検出手段と、(c) 画像データを入力する画像入力手段と、(d) 入力した画像データを所定の捕捉間隔で捕捉する画像捕捉手段と、(e) 画像データを所定の圧縮率で符号化圧縮する画像符号化圧縮手段と、(f) 前記音声検出手段による検出結果に応じて、音声データを符号化圧縮するとともに、符号化圧縮された画像データと混合して、パケットを生成するデータ混合手段と、(g) 生成したパケットを伝送する伝送手段と、(h) 前記音声検出手段による過去1回以上の検出結果を基に、以後の音声データの有無を予測する音声予測手段と、(i) 音声予測手段による予測結果に応じて、前記画像捕捉手段の前記所定の捕捉間隔を調整する調整手段と、を具備することを特徴とするデータ伝送装置である。

【0030】ここで、前記画像捕捉手段は比較的短い第1の捕捉間隔又は比較的低速な第2の捕捉間隔のうちのいずれか一方で画像データを捕捉し、また、前記調整手段は、予測結果が音声データ有りを示す場合は第2の捕捉間隔を選択し、予測結果が音声データ無しを示す場合は第1の捕捉間隔を選択する、ようにしてもよい。

【0031】また、本発明の第6の側面は、符号化圧縮した音声データと符号化圧縮した画像データを固定ビット幅のパケットに混合してネットワーク上に伝送するタイプのデータ伝送装置において、(a) 音声データの入力を試みる音声入力手段と、(b) 音声データの入力の有無を検出する音声検出手段と、(c) 画像データを入力する画像入力手段と、(d) 入力した画像データを所定の捕捉間隔で捕捉する画像捕捉手段と、(e) 画像データを所定の圧縮率で符号化圧縮する画像符号化圧縮手段と、(f) 前記音声検出手段による検出結果に応じ

て、音声データを符号化圧縮するとともに、符号化圧縮された画像データと混合して、パケットを生成するデータ混合手段と、(g)生成したパケットを伝送する伝送手段と、(h)前記音声検出手段による過去1回以上の検出結果を基に、以後の音声データの有無を予測する音声予測手段と、(i)音声予測手段による予測結果に応じて、前記画像符号化圧縮手段の前記所定の圧縮率を調整する調整手段と、を具備することを特徴とするデータ伝送装置である。

【0032】ここで、前記画像符号化圧縮手段は比較的高い第1の圧縮率又は比較的低い第2の圧縮率のうちのいずれか一方で画像データを捕捉し、また、前記調整手段は、予測結果が音声データ有りを示す場合は第1の圧縮率を選択し、予測結果が音声データ無しを示す場合は第2の圧縮率を選択する、ようにしてもよい。

【0033】

【作用】音声データと画像データとを混合して固定ビット幅のパケットを生成する場合、プライオリティを与えられた音声データが存在するかどうかによって、画像データに割り当てられる帯域幅を大いに相違する。例えば図10に示す例では、画像データの帯域幅は16ビット又は286ビットであり、その差は非常に大きい。

【0034】本発明に係るデータ伝送方法及び装置によれば、音声データが無いと考えられ、したがって、比較的広い帯域幅が画像データに与えられる期間には、画質を重視してパケットを生成する。すなわち、画像データの圧縮率を低くし、あるいは画像データをキャプチャリングするフレーム・レートを上げるなどして、許容される範囲で、画質の向上を図っている。

【0035】逆に、音声データが有ると考えられ、したがって、比較的狭い帯域幅しか画像データに与えられない期間には、画像データのトラフィックを重視してパケットを生成する。すなわち、画像データの圧縮率を高くし、あるいは、画像データをキャプチャリングするフレーム・レートを下げるなどして、データの滞り(1フレーム以上の無意味なバッファリング)を生じないようにしている。

【0036】したがって、本発明によれば、符号化圧縮した音声データと符号化圧縮した画像データを多重化して固定ビット幅長のパケットにして伝送するときに、音声データにプライオリティを与えながらも画像データも滞りなく伝送することができる。

【0037】また、本発明によれば、符号化圧縮した画像データを固定ビット幅のパケット内に入れて伝送する場合において、画像データに割り当てられる帯域幅が変動しても、好適に画像データを伝送することができる。換言すれば、本発明によれば、滞らず且つ隙間なく、画像データを伝送できる訳である。

【0038】本発明のさらに他の目的、特徴や利点は、後述する本発明の実施例や添付する図面に基づくより詳

細な説明によって明らかになるであろう。

【0039】

【発明の実施の形態】以下、図面を参照しながら本発明の実施例を詳解する。

【0040】A. コンピュータ・システムのハードウェア構成

図1には、本発明の実施に供されるコンピュータ・システム100のハードウェア構成を模式的に示している。システム100は、一般公衆回線上に接続された1つのコンピュータ・システムに該当する。以下、各部について説明する。

【0041】メイン・コントローラであるCPU11は、オペレーティング・システム(OS)の制御下で、各種プログラムを実行するようになっている。CPU11の動作クロックは、発振器(OSC)12から供給される。CPU11は、例えば、米IBM社、米Motorola社及び米Apple社が共同開発した"PowerPC 603e-100MHz"("PowerPC"は米IBM社の商標)でよい。CPU11は、自身の外部ピンに直結したプロセッサ・バス16、ローカル・バスとしてのPCI(Peripheral Component Interconnect)バス17、及び入出力バスとしてのISA(Industry Standard Architecture)バス23という3階層のバスを介して、各デバイスと相互接続している。

【0042】プロセッサ・バス16とPCIバス17とは、ブリッジ回路(ホスト-PCIブリッジ)13によって連絡されている。本実施例のブリッジ回路13は、メイン・メモリ15へのアクセス動作を制御するためのメモリ・コントローラと、両バス16、17間の速度差を吸収するためのデータ・バッファを含んだ構成となっている。メイン・メモリ15は、DRAMのような書き込み可能な半導体メモリで構成され、各プログラムの格納領域、及び実行中のプログラムの作業領域として用いられる。メイン・メモリ15の記憶容量は、通常、数MB~数十MB程度である。また、L2-キャッシュ14は、SRAMのような高速アクセス可能な半導体メモリで構成され、CPU11の処理速度とメモリ15へのアクセス速度のギャップを吸収するために、必要最小限のデータの一時的保持に用いられる。L2-キャッシュ14の記憶容量は、例えば256KBである。また、ROM18は、ハードウェア操作のための制御コード(BIOS)や、始動時のテスト・プログラム(POST)などを恒久的に記憶するための不揮発性半導体メモリである。

【0043】PCIバス17は、米Intel社の提唱に基づいて規格化されたバスであり、バス幅32ビット、動作周波数33MHz、最高データ転送速度132Mbpsを主な性能としている。PCIバス17には、グラフィック・コントローラ19やビデオ・キャプチャ・コントローラ20などのような、比較的高速なデータ



転送を要するP C Iデバイスが接続される。

【0044】グラフィック・コントローラ19は、コンピュータ画像の表示処理を行うための周辺コントローラであり、CPU11からの描画命令に従って描画情報を画面バッファ(VRAM)19bに一旦書き込むとともに、VRAM19bから描画情報を読み出して、標準装備された液晶表示ディスプレイ(LCD)19aに出力するようになっている。また、グラフィック・コントローラ19は、読み出した描画情報を付設するDAコンバータ19cによってアナログ化して、外付けCRT(Cathod Ray Tube)ディスプレイ19dに出力することもできる。

【0045】ビデオ・キャプチャ・コントローラ20は、ビデオ・カメラ20a(又はVTR:図示しない)からアナログ入力した映像信号をデジタル化してファイルに落とすためのコントローラである。ビデオ・キャプチャ・コントローラ20は、一般には、ビデオ・キャプチャ用デバイス・ドライバ(後述)によってハードウェア操作され、フレーム(すなわち画面)単位でキャプチャリングを行うようになっている。フレーム・レート(すなわち単位時間当りにキャプチャリングするフレーム数)は、ハードウェアの許容する範囲(通常は15~30フレーム/秒程度)で制御可能である。ビデオ・キャプチャ・コントローラ20は、ビデオ会議のための画像データの捕捉にも用いられる。

【0046】P C Iバス17とI S Aバス23とは、ブリッジ回路(P C I-I S Aブリッジ)21によって相互接続されている。本実施例のブリッジ回路21は、DMAコントローラ、プログラマブル割込みコントローラ(PIC)、プログラマブル・インターバル・タイマ(PIT)を含んだ構成となっている。さらに、ブリッジ回路21は、ハード・ディスク・ドライブ(HDD)22を接続するためのI D Eインターフェース(Integrated Drive Electronics:I D Eは本来I S AバスにHDDを直結させるための規格)を含んでいる。

【0047】I S Aバス23上には、キーボード/マウス・コントローラ(KMC)24、シリアルI/Oコントローラ25、パラレルI/Oコントローラ26、フロッピー・ディスク・コントローラ(FDC)27a、オーディオ・コントローラ28などの、比較的低速なデータ転送で済むI S Aデバイスが接続されている。

【0048】KMC24は、キーボード24aからの入力マトリックスやマウス24bによる指示座標値を処理するためのコントローラである。

【0049】シリアルI/Oコントローラ25は、シリアル・ポート25aを介して行われる、他の機器との間のシリアル・データ転送を制御するためのものである。シリアル・ポート25aには例えばモデムが装着される。モデムは、一般公衆回線のようなアナログ通信回線を介してデータ通信を行うための信号変換装置である。

換言すれば、コンピュータ・システム100は、モデムによって、一般公衆回線を利用したビデオ会議に参入することができる訳である。モデムの最大転送レートは、例えば14.4k b p s(又は28.8k b p s)程度である。

【0050】パラレルI/Oコントローラ26は、パラレル・ポート26aを介して行われる、他の機器との間のパラレル・データ転送を制御するためのものである。パラレル・ポート26aに接続される装置の代表例は、プリンタである。

【0051】FDC27aは、フロッピー・ディスク・ドライブ27の駆動を制御するためのコントローラである。

【0052】オーディオ・コントローラ28は、オーディオ・ラインイン・ジャック28aに接続されたマイクからのオーディオ入力や、オーディオ・ラインアウト・ジャック28bに接続されたスピーカによるオーディオ出力を制御するためのコントローラである。オーディオ・コントローラ28は、オーディオ用デバイス・ドライバ(後述)によってハードウェア操作される。ビデオ会議のための音声データの入出力は、オーディオ・コントローラによって行われる。

【0053】コンピュータ・システム100は、デスクトップP C、ノートブックP C、あるいはその他のハイエンド・マシンであってもよい。システム100の一例は、日本アイ・ビー・エム(株)が市販する"IBM ThinkPad Power Series 850"("ThinkPad"は米IBM社の商標)である。

【0054】なお、コンピュータ・システム100を構成するためには、図1に示した以外にも多くの電気回路等が必要であるが、これらは当業者には周知であり、且つ本発明の要旨とは関連がないので、本明細書中では省略してある。

## 【0055】B. コンピュータ・システム上で稼働するソフトウェア

### B-1. ソフトウェアの階層的構成

図2には、コンピュータ・システム100で実行される各ソフトウェア間の連携的構成を模式的に示している。

【0056】最下層のソフトウェアは、デバイス・ドライバ(D/D)のような、ハードウェアを直接制御するためのプログラムである。例えば、グラフィック・コントローラ19を操作するためのディスプレイ用デバイス・ドライバや、ビデオ・キャプチャ・コントローラ20を操作するためのビデオ・キャプチャ用デバイス・ドライバ、シリアルI/Oコントローラ25を操作するためのCOMデバイス・ドライバ、オーディオ・コントローラ28を操作するためのオーディオ用デバイス・ドライバなどがこれに該当する。

【0057】オペレーティング・システム(OS)は、

システム100のハードウェア及びソフトウェアを総合的に管理するための基本ソフトウェアであり、例えば、OS/2("OS/2"は米IBM社の商標)がこれに該当する。OSは、HDD22などの記憶装置に格納されたファイルを管理するための「ファイル・マネージャ」、メモリ領域の割り当てを管理するための「メモリ・マネージャ」、CPU11のタスク実行の順序を管理するための「スケジューラ」を含んでいる。また、ウィンドウ表示やキーボード/マウス操作等の処理のための「ユーザ・インターフェース」も含んでいる。

【0058】OS層の上位には、OSの管理下で実行される各種アプリケーション・プログラムが存在する。アプリケーション・プログラムは、必要に応じて適宜HDD22などの外部記憶装置からメモリ15にロードされる。本発明の具現に関連するアプリケーション・プログラムは、ビデオ会議制御プログラム、ビデオ・アプリケーション・プログラム、データ混合分離プログラムである。以下、各アプリケーション・プログラムについて簡単に説明する。

【0059】ビデオ会議制御プログラム：ビデオ会議制御プログラムは、会議(セッション)の開始・終了(電話をかける・切るなどの操作)、ディスプレイ画面上のビデオ表示の開始・終了、ビデオ・アプリケーションによる圧縮作業の開始・終了、音声の音量やマイクロフォンのゲイン調整など、会議全体の制御を行うためのソフトウェアである。モデムの接続速度の表示やエラー・メッセージの表示も行う。

【0060】ビデオ・アプリケーション・プログラム：ビデオ・アプリケーション・プログラムは、画像データの捕捉やパケット転送に伴う以下の機能を行うためのソフトウェアである。

- 1) ビデオ・キャプチャ・コントローラ20によって取り込まれた画像フレームを符号化圧縮する。
- 2) 符号化圧縮された画像フレームを、データ混合分離プログラムに渡す。
- 3) データ混合分離プログラムより渡された画像フレームを復号化伸長する。
- 4) 復号化伸長された画像フレームを、グラフィック・コントローラ19に渡して、コンピュータ画面上で再生する(但し、グラフィック・コントローラ19へのハードウェア的な入出力動作には、ディスプレイ・デバイス・ドライバが介在する。)

【0061】ビデオ・アプリケーション・プログラムは、画像データを圧縮・伸長処理するための画像符号化圧縮/復号化伸長モジュールを含んでいる。画像圧縮/伸長モジュールは、例えば「MPEG1」に従うものでよい。画像圧縮/伸長モジュールは、一般には、フレーム単位で画像データの圧縮及び伸長処理を行うようになっている。したがって、送信側では、1フレームの圧縮を完了する前に取り込まれた次のフレームの画像データは、

バッファリングされ、データの停滞の原因になる。また、受信側では、1フレーム分の画像データを受け取るまでは画像データの復号化伸長を行えないので、1フレーム分の画像データが膨大で転送時間が長引けば、表示処理の遅延につながる。また、画像符号化圧縮モジュールは、与えられたパラメータ(例えばビット・レート)に応じて、フレーム・レートを維持するように、すなわち一定間隔で画像フレームを処理するように稼働する。例えば画像データの送信が滞っている場合には、圧縮率を上げる(すなわち画質を下げる)ことによってフレーム・レートの維持に努める。逆に画像データの送信に余裕がある場合には、圧縮率を下げる(すなわち画質を上げる)ようになっている。

【0062】データ混合分離プログラム：データ混合分離プログラムは、音声データと画像データを混合してパケットを生成したり、受け取ったパケットから音声データや画像データを組み立てるためのソフトウェアである。主な機能は以下の通りである。

- 1) オーディオ・コントローラ28により入力した音声データを、符号化圧縮する(但し、オーディオ・コントローラ28へのハードウェア的な入出力動作には、オーディオ・デバイス・ドライバが介在する。)
- 2) ビデオ・アプリケーションから渡された符号化圧縮済みの画像データと、自身で符号化圧縮した音声データとを、混合する。
- 3) 混合したデータをパケットに分割する。
- 4) パケットをシリアルI/Oコントローラ25に渡す(但し、シリアルI/Oコントローラ25へのハードウェア的な入出力動作には、COMデバイス・ドライバが介在する)。シリアルI/Oコントローラ25に渡されたパケットは、モデム、一般公衆回線経由で受信側のコンピュータ・システムに伝送される。
- 5) 受信側のコンピュータ・システムから伝送されてきたパケットを、シリアルI/Oコントローラ25より受け取る(但し、シリアルI/Oコントローラ25へのハードウェア的な入出力動作には、COMデバイス・ドライバが介在する。)
- 6) 受け取ったパケットを組み立てる。
- 7) 組み立てられたデータを、音声データと画像データとに分離する(但し、音声データと画像データは、ともに符号化圧縮されたままの状態)。
- 8) 画像データを、ビデオ・アプリケーションに渡す。
- 9) 音声データを復号化伸長して、オーディオ・コントローラ28により再生する(但し、オーディオ・コントローラ28へのハードウェア的な入出力動作には、オーディオ・デバイス・ドライバが介在する。)
- 10) モデムのコントロール、音声の入出力の一時停止、再開、接続後の回線の状況の監視を行う。

【0063】データ混合分離プログラムは、現在の音声の入力の有無(すなわちVoice Activity)

も監視して、この監視結果をVoice Activityビットに書き込むようになっている。パケットのGSMビットは、Voice Activityに応じて設定／解除するとともに、パケットの構造を決める（図10参照）。本実施例では、データ混合分離プログラムはVoice Activityビットを20ms毎に更新するようになっている。また、データ混合分離プログラムをライブラリ形式にしておくことにより、これとリンクした他のアプリケーション・プログラムと同一のプロセス空間内で動作し、同じリソースとしてVoice Activityビットを共有することができる。すなわち、アプリケーション・プログラム（例えばビデオ・アプリケーション）はVoice Activityビットを自ら直接参照することができる訳である。

【0064】なお、OS層とアプリケーション層の間に、破線で図示したようなプラットフォーム層が介在していてもよい。プラットフォーム層は、通信回線によって相互接続されたコンピュータ・システム間で、ソフトウェアやデータのシェアリングを可能にするためのものである。プラットフォーム層は“Collaboration Framework”と呼ばれることもある。

#### 【0065】B-2. 各ソフトウェア間の協働的オペレーション

ここで、各ソフトウェア間の協働的動作を、簡単に説明しておく。

【0066】図3には、入力した音声データと画像データを分割してパケット化して一般公衆回線上に伝送する様子を、模式的に示している。

【0067】マイクを介して入力された音声データは、オーディオ・デバイス・ドライバによる入出力操作によって、データ混合分離プログラムに渡される（矢印P1, P2, P3）。データ混合分離プログラムでは、音声データはGSMアルゴリズムなどに従って、符号化圧縮される。

【0068】一方、ビデオ・カメラ20aによって撮像された画像フレームは、ビデオ・キャプチャ・コントローラ20によってデジタル化される（矢印P4, P5）。ビデオ・アプリケーションは、ビデオ・キャプチャ・デバイス・ドライバによる入出力操作によって、1フレーム単位で画像データを取得し（矢印P6）、これをMPEG1などに従って符号化圧縮する。フレーム・レートが速過ぎるなどの理由により、1フレームを符号化圧縮処理する前に次の1フレームをキャプチャした場合には、次の1フレーム分の画像データはバッファリングされ、データの停滞の原因になる。

【0069】データ混合分離プログラムは、ビデオ・アプリケーションより符号化圧縮された画像データを受け取り（矢印P7）、符号化圧縮された音声データと画像データとを混合し、混合データを所定ビット幅のパケッ

トに分割する。このとき、Voice Activityビットの内容に応じて、ヘッダ部のGSMビットが設定／解除されるとともに、パケット構造が決定される。音声データは、圧縮方式で定められた所定の帯域幅（本実施例では264ビット幅の帯域）を優先して占有する。また、画像データは、その余の帯域（音声有りの場合は16ビット幅の帯域、音声無しの場合は286ビット幅の帯域）に入れられる（図10参照）。フレーム・サイズが膨大な画像データを伝送する場合や、音声有りの状態が継続する場合には、所要パケット数が多くなり、画像1フレーム分の伝送時間が長くなることになる。

【0070】逐次分割されたパケットは、COMデバイス・ドライバの入出力操作によって、シリアルI/Oコントローラ25に渡される（矢印P9, P10）。シリアルI/Oコントローラ25は、モデム経由で一般公衆回線上にパケットを伝送する。

【0071】また、図4には、一般公衆回線を介して受け取ったパケットから音声データと画像データを組み立てる様子を、模式的に示している。

【0072】一般公衆回線を経由して送られてきたパケットは、モデムによって受け取られ且つデジタル化される。データ混合分離プログラムは、シリアルI/Oデバイス・ドライバによる入出力操作によって、パケットを受け取る（矢印P11, P12, P13）。

【0073】データ混合分離プログラムは、パケットを組み立てるとともに、音声データと画像データに分離する。音声データは、データ混合分離プログラム内で復号化伸長処理され、オーディオ・コントローラ28に渡される（矢印P13, P14）。オーディオ・コントローラ28は、音声データをスピーカによって音声出力する（矢印P15）。

【0074】一方、画像データは、ビデオ・アプリケーションに渡される（矢印P16）。ビデオ・アプリケーションは、画像データを復号化伸長して、1フレーム分の画像データを組み立てる。

【0075】組み立てられた画像フレームは、ディスプレイ・デバイス・ドライバによる入出力操作によってグラフィック・コントローラ19に渡され（矢印P17, P18）、LCD19b画面上に表示される（矢印P19）。なお、1フレームのデータ・サイズが膨大である、あるいはパケット中で画像データに割り当てられた帯域幅が狭いなどの理由により、画像1フレーム分のデータ伝送に長時間を要した場合には、ビデオ・アプリケーションによる画像フレームの出力は遅延する。この結果、受信側のLCD19b画面上には数秒前の画像が表示されることになる。

#### 【0076】C. 画像データ伝送の最適化オペレーション

本発明の特徴を略言すれば、現在までの音声データの有

無に従って、将来通信路中で画像データに割り当てられている帯域幅を予知し、画像データの伝送の最適化を図ることである。音声データの有無は、データ混合分離プログラムが返すVoice Activityビットを参照することによって判定できる。過去所定期間内のVoice Activityビットの履歴を基にすれば、将来音声入力があるかどうかを経験的に予測することができる。この予測結果は将来画像データに割り当てられるべき帯域幅を一義的に意味する。この予測された帯域幅に応じて画像データ伝送の最適化を図ればよい、という訳である。

【0077】ここで言う画像データ伝送の最適化とは、データが滞らず、且つ隙間なく伝送することを意味する。最適化処理の具体的な例として、(1) 音声の予測結果を符号化圧縮工程にフィードバックする手法と、

(2) 音声の予測結果を画像捕捉工程にフィードバックする手法とが挙げられる。前者によれば、予測された帯域幅に応じて画像データの圧縮率を調整することによって、データ伝送の遅延を最小化することができる。また、後者によれば、予測された帯域幅に応じて次の画像フレームをキャプチャすべき時間間隔を設定することにより、隙間なく画像フレームを供給することができる。

#### 【0078】C-1. 遅延の最小化

この項では、音声の予測結果を符号化圧縮工程にフィードバックする手法について説明する。図5には、同手法の概念図を示している。

【0079】この例のビデオ・アプリケーションは、画像データをフレーム単位で符号化圧縮するための画像符号化圧縮モジュールの他に、キャプチャ・モジュールと、音声予測モジュールを含んでいる。キャプチャ・モ

(ビット・レート)

$$=(\text{モデムDCE速度})-(\text{オーバーヘッド})-(\text{音声データのビット・レート})$$

…(1)

【0084】ここで、モデムDCE(Data Communication Equipment)速度とは、モデムの最大転送速度と等価であり、例えば14.4kbpsである。モデムDCE速度は、データ混合分離プログラムにより設定され、一旦回線に接続された後は固定的に用いられる。また、オーバーヘッドは、パケットのヘッダ部に割かれるビット・レートであり、固定値である(図10(a)では0.4kbps)。また、音声データのビット・レートは、優先して割かれる固定値であり、GSMアルゴリズムを用いる場合(図10(a)参照)には、13.2k

$$(\text{ビット・レート})=(\text{モデムDCE速度})-(\text{オーバーヘッド}) \quad \dots(2)$$

【0087】この場合、オーバーヘッド(0.1kbps (=2ビット÷20msec))以外の全てのモデムDCE速度、すなわち14.3kbpsが画像符号化圧縮モジュールにフィードバックされることになる。

【0088】画像符号化圧縮モジュールは、新たに受け

ジュールは、ビデオ・キャプチャ・コントローラ20に対して、1フレーム分の画像データの供給を要求する機能を備えている。また、音声予測モジュールは、過去の所定期間内の音声入力の履歴に従ってこれから音声入力があるかどうかを予測し、予測結果に基づいて、最適なビット・レートを画像符号化圧縮モジュールにフィードバックする機構を備えている。以下、データの流れに従って、各部の動作について説明する。

【0080】第(N-1)フレーム目の画像データが入力されると、画像符号化圧縮モジュールは、これを符号化圧縮してデータ混合分離プログラムに渡す。データ混合分離プログラムは、この時点で入力した音声データを符号化圧縮するとともに、渡された画像データと混合し、パケットに分割する。さらに、データ混合分離プログラムは、音声入力の有無を示すVoice Activityビットを20msec毎に更新する。

【0081】音声予測モジュールは、所定の時間周期t[sec]ごとにVoice Activityビットを読み、且つ所定時間T[sec](但しT>t)内におけるVoice Activityビットの履歴を参照し、所定の音声予測アルゴリズム(D項参照)に従って、次の第Nフレーム目の画像データを転送する間に音声入力があるかどうかを予測する。

【0082】さらに、音声予測モジュールは、予測結果に基づいて最適なビット・レートを計算して、画像符号化圧縮モジュールにフィードバックする。音声予測が“True”、すなわち音声入力があると予測された場合、最適なビット・レートは以下の式(1)により算出される。

【0083】

【数1】

bps(=(260+4)ビット÷20msec)である。したがって、画像符号化圧縮モジュールにフィードバックされるビット・レートは0.8kbpsということになる。

【0085】一方、音声予測が“False”、すなわち音声入力がないと予測された場合、最適なビット・レートは以下の式(2)により算出される。

【0086】

【数2】

取ったビット・レートに応じて、フレーム・レートを維持するように、第Nフレーム目の画像データの符号化圧縮処理を行う。以下、同様の処理が繰り返されることになる。

【0089】この方法によれば、音声予測がTrueで

あれば必然的に圧縮率が上がり(画質が低下し)、Falseであれば圧縮率が下がる(画質が向上する)。この結果、データ伝送の遅延を最小化することができる。

【0090】なお、音声入力の有無を予測するためのアルゴリズムについては、次項Dで詳解する。

#### 【0091】C-2. フレーム・レートの最大化

この項では、音声の予測結果を画像捕捉工程にフィードバックする手法について説明する。図6には、同手法の概念図を示している。

【0092】この例のビデオ・アプリケーションは、画像データをフレーム単位で符号化圧縮するための画像符号化圧縮モジュールの他に、キャプチャ・モジュールと、音声予測モジュールを含んでいる。キャプチャ・モジュールは、ビデオ・キャプチャ・コントローラ20に対して、1フレーム分の画像データの供給を要求する機能を備えている。また、音声予測モジュールは、過去の所定期間内の音声入力の履歴に従ってこれから音声入力があるかどうかを予測し、予測結果に基づいて、次の画像フレームを取り込むべき最適な時間間隔をキャプチャ・モジュールにフィードバックする機構を備えている。以下、データの流れに従って、各部の動作について説明する。

【0093】キャプチャ・モジュールが発行した入力要求に応じて、第(N-1)フレーム目の画像データが入力される。画像符号化圧縮モジュールは、これを符号化圧縮してデータ混合分離プログラムに渡す。データ混合分離プログラムは、この時点で入力した音声データを符号化圧縮するとともに、渡された画像データと混合し、パケットに分割する。さらに、データ混合分離プログラムは、音声入力の有無を示すVoice Activityビットを20msec毎に更新する。

【0094】音声予測モジュールは、所定の時間周期t[sec]ごとにVoice Activityビットを読み、且つ所定時間T[sec](但しT>t)内におけるVoice Activityビットの履歴を参照し、所定のアルゴリズムに従って、次の第Nフレーム目の画像データを転送する間に音声入力があるかどうかを予測する。

【0095】さらに、音声予測モジュールは、予測結果に基づいて、第(N-1)フレームの画像データを取り込んでから第Nフレーム目の画像データを取り込むまでの最適な時間間隔T<sub>c</sub>を計算して、キャプチャ・モジュールにフィードバックする。ここで言う最適な時間間隔T<sub>c</sub>とは、データの滞りも隙間もなく画像フレームを取り込めるタイミングを意味する。したがって、時間間隔T<sub>c</sub>は以下の式(3)により算出される。

【0096】

【数3】

$$T_c = F_r \times 10 / B + \alpha \quad \cdots (3)$$

【0097】ここで、F<sub>r</sub>は、第(N-1)フレーム目

の符号化圧縮後のフレーム・サイズ(バイト数)であり、Bはパケット中で画像データに割り当てられた帯域幅である。第1項で10を乗算しているのは、1バイト(=8ビット)分のデータをシリアル転送する場合には、その前後各端に同期用のスタート/ストップ・ビットを足した10ビット長にして伝送するからである。式(3)の右辺第1項は、第(N-1)フレーム目の画像データの伝送に要する時間に該当する。また、同第2項の定数αは、画像フレームを取り込んだ後の処理(例えば画像符号化圧縮処理やデータ混合処理)に要する時間であり、CPU11に依存する値である。

【0098】キャプチャ・モジュールは、T<sub>c</sub>秒後に、第Nフレーム目の画像データの入力を要求する。以下、同様の処理が繰り返されることになる。

【0099】この方法によれば、画像フレームの伝送が完了する前に次の画像フレームの取り込みを開始して無益なデータ・バッファリングを誘発する、ということはない。また、許容範囲で最大限の画像フレームの取り込みに努めるため、既に画像フレームの伝送が完了しているにも拘らず次の画像フレームの取り込みを行わない、という帯域の浪費も避けることができる。

【0100】なお、音声入力の有無を予測するためのアルゴリズムは、C-1項と同様のものでよい。詳しくは次項Dで説明する。

#### 【0101】D. 音声予測アルゴリズム

C項で説明したように、音声入力の予測結果に基づいてデータ伝送の最適化が図られるようになっている。この項では、図5及び図6の音声予測モジュールに適用可能な音声予測アルゴリズムについて説明する。

【0102】図7には、音声予測アルゴリズムの第1の例を、フローチャート化して示している。該アルゴリズムは、データ混合分離プログラムより取得した現在及び過去一定期間のVoice Activityビットに従って、将来の音声入力の有無を予測する、ということの基本原則としている。

【0103】音声予測モジュールは、タイマー機能を持っており、t[sec]周期でVoice Activityビットを参照し、その値(すなわちTrueかFalseか)を自身のバッファ(以下、「音声予測バッファ」という)に逐次書き込む(ステップS10)。この音声予測バッファは、T[sec](T>t)分、すなわち過去複数個のVoice Activityビットの値を書き込む記憶容量を持ち、バッファが一杯になると古い順にデータを廃棄して新しいデータを書き込む、という性質を持っている。

【0104】次いで、音声予測モジュールは、音声予測バッファを参照することによって、最近T[sec]間の音声入力の履歴を調べる。本例では、過去T[sec]間でのTrueの割合を調べ(ステップS20)、所定値P[%]を上回れば、画像符号化圧縮モジュール

(又はキャプチャ・モジュール)に対してTrueを出力し(ステップS30)、P[%]未満であればFalseを出力する(ステップS40)。

【0105】図8には、音声予測アルゴリズムの第2の例を、フローチャート化して示している。該アルゴリズムも、図7と同様に、現在及び過去一定期間のVoice Activityビットに従って、将来の音声入力の有無を予測することを基本原理としている。

【0106】音声予測モジュールは、タイマー機能を持っており、t[sec]周期でVoice Activityビットを参照し、その値(すなわちTrueかFalseか)を自身のバッファ(以下、「音声予測バッファ」という)に逐次書き込む(ステップS110)。この音声予測バッファは、T[sec](T>t)分、すなわち過去複数個のVoice Activityビットの値を書き込む記憶容量を持ち、バッファが一杯になると古い順にデータを廃棄して新しいデータを書き込む、という性質を持っている。

【0107】次いで、音声予測モジュールは、音声予測バッファを参照することによって、最近T[sec]間の音声入力の履歴を調べる。本例では、過去T[sec]間におけるVoice Activityビットの時間による加重平均値 $A_w$ を計算するようになっている(ステップS120)。加重平均値 $A_w$ は、下式(4)により求められる。

【0108】

【数4】

$$A_w = \frac{1}{N} \cdot \sum_{i=0}^N V_i W_i \quad \dots (4)$$

【0109】ここで、Nは、音声予測バッファ内のデータの個数である。また、 $V_i$ は音声予測バッファ内のi番目のVoice Activityビットの値であり、Trueであれば1とし、Falseであれば0とする。また、 $W_i$ はi番目のデータに対する時間的な重み因子であり、新しいデータほど大きく評価するためである。

【0110】 $A_w$ が閾値Qを上回れば、画像符号化圧縮モジュール(又はキャプチャ・モジュール)に対してTrueを出力し(ステップS130)、閾値Q未満であればFalseを出力する(ステップS140)。

【0111】なお、所定時間T[sec]だけ遡って音声予測するのは、人間は一度話し始めるとある程度の期間は話し続ける、という習性に依拠する。Voice Activityビットは20msec毎の音声入力の有無を示すに過ぎない。これに対して、ある一定期間T[sec]だけ該ビットを監視した方が、1フレーム分の画像データを伝送する間(通常複数パケットに跨がる)という、比較的長期の音声予測をより正確に行うことができる。

## 【0112】E. 追補

以上、特定の実施例を参照しながら、本発明について詳解してきた。しかしながら、本発明の要旨を逸脱しない範囲で当業者が該実施例の修正や代用を成し得ることは自明である。すなわち、例示という形態で本発明を開示してきたのであり、限定的に解釈されるべきではない。本発明の要旨を判断するためには、冒頭に記載した特許請求の範囲の欄を参酌すべきである。

## 【0113】

【発明の効果】本発明に係るデータ伝送方法及び装置は、音声データが無いと考えられ、したがって、比較的広い帯域幅が画像データに与えられる期間には、画質を重視してパケットを生成する。すなわち、画像データの圧縮率を低くし、あるいは画像データをキャプチャリングするフレーム・レートを上げるなどして、許容される範囲で、画質の向上を図っている。

【0114】逆に、音声データが有ると考えられ、したがって、比較的狭い帯域幅しか画像データに与えられない期間には、画像データのトラフィックを重視してパケットを生成する。すなわち、画像データの圧縮率を高くし、あるいは、画像データをキャプチャリングするフレーム・レートを下げるなどして、データの滞り(1フレーム以上の無意味なバッファリング)を生じないようにしている。

【0115】したがって、本発明によれば、符号化圧縮した音声データと符号化圧縮した画像データを多重化して固定長のパケットにして伝送するときに、音声データにプライオリティを与えながらも画像データも滞りなく伝送することができる、優れたデータ伝送方法及び装置を提供することができる。

【0116】また、本発明によれば、符号化圧縮した画像データを固定長のパケット内に入れて伝送する場合において、画像データに割り当てられる帯域幅が変動しても、好適に画像データを伝送することができる優れたデータ伝送方法及び装置を提供することができる。換言すれば、本発明によれば、映像の遅延と品質を最適化して、画像データを伝送できる訳である。

## 【図面の簡単な説明】

【図1】図1は、本発明の実施に供されるコンピュータ・システム100のハードウェア構成を模式的に示した図である。

【図2】図2は、コンピュータ・システム100で実行される各ソフトウェアの構成を模式的に示した図である。

【図3】図3は、入力した音声データと画像データを分割してパケット化して一般公衆回線に伝送する様子を、模式的に示した図である。

【図4】図4は、一般公衆回線を介して受け取ったパケットから音声データと画像データを組み立てる様子を、模式的に示した図である。

【図5】図5は、音声の予測結果を符号化圧縮工程にフィードバックする手法の概念図を示している。

【図6】図6は、音声の予測結果を画像捕捉工程にフィードバックする手法の概念図を示している。

【図7】図7は、音声予測アルゴリズムの第1の例をフローチャート化して示した図である。

【図8】図8は、音声予測アルゴリズムの第1の例をフローチャート化して示した図である。

【図9】図9は、一般公衆回線とPCを用いたネットワークの構成を模式的に示した図である。

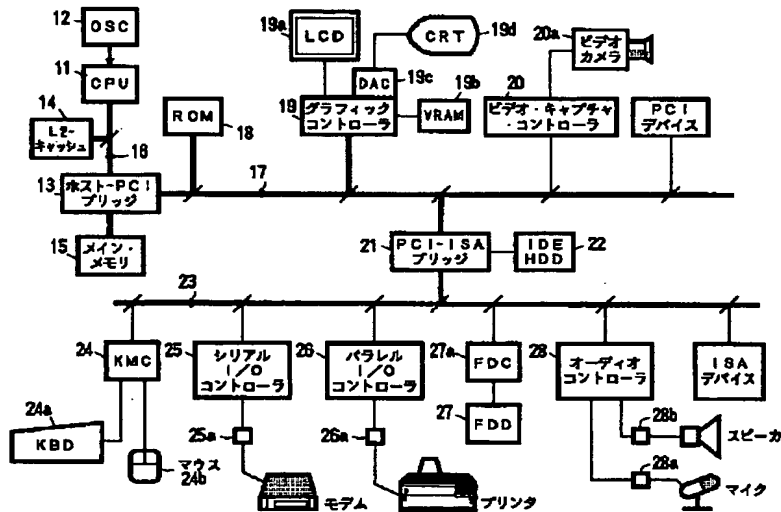
【図10】図10は、音声データ及び画像データを伝送

するためのバケット構造の一例を示した図である。

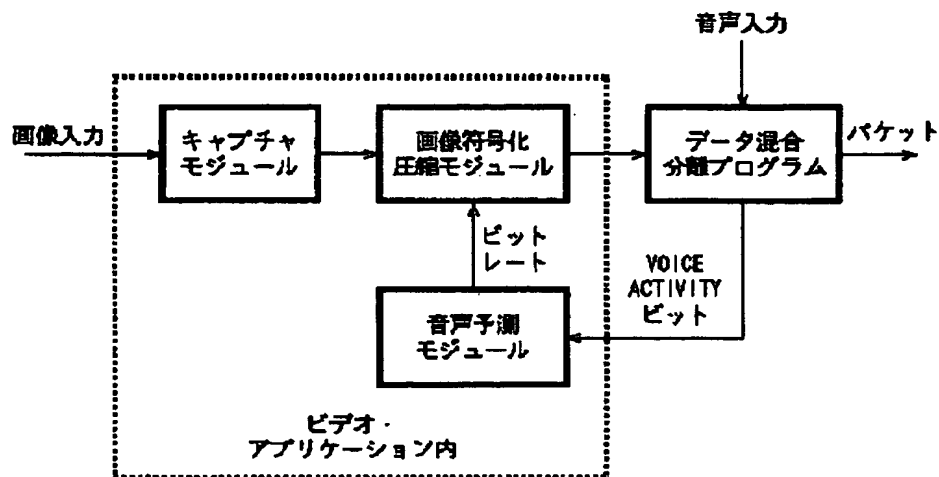
【符号の説明】

11…CPU、12…発振器、13…ブリッジ回路、14…L2-キャッシュ、15…メイン・メモリ、16…プロセッサ・バス、17…PCIバス、18…ROM、19…グラフィック・コントローラ、20…ビデオ・キャプチャ・コントローラ、21…ブリッジ回路、22…HDD、23…ISAバス、24…KMC、25…シリアルI/Oコントローラ、26…パラレルI/Oコントローラ、27…FDD、28…オーディオ・コントローラ、100…コンピュータ・システム。

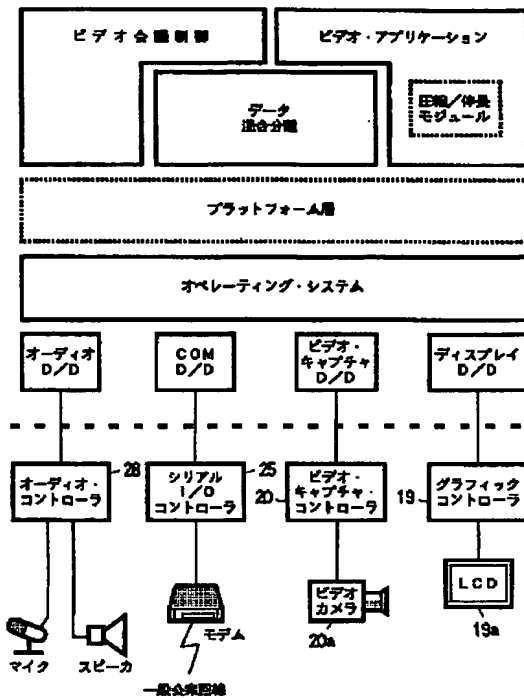
【図1】



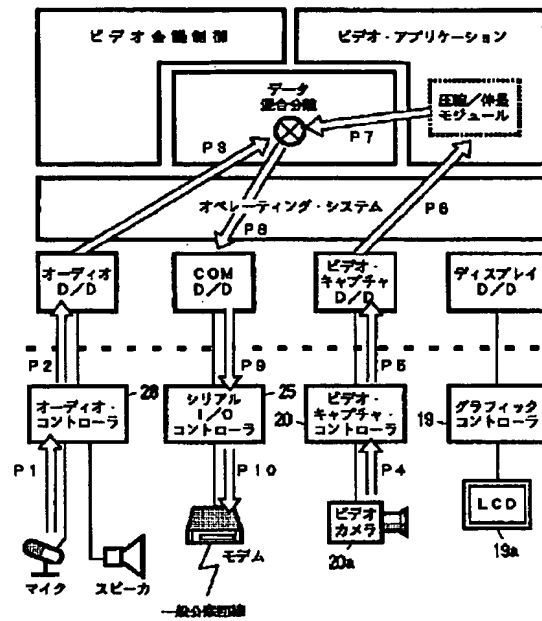
【図5】



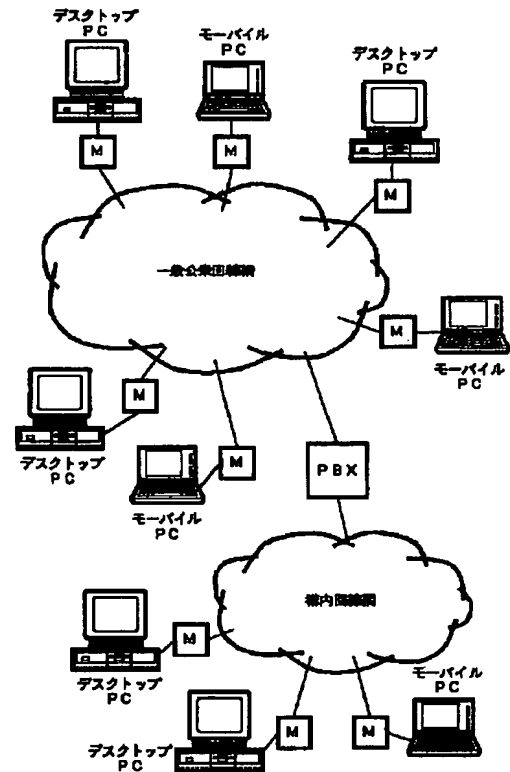
【図2】



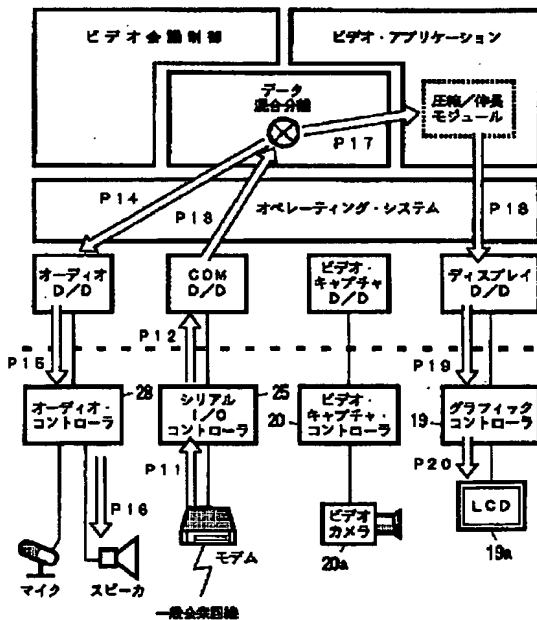
【図3】



【図9】

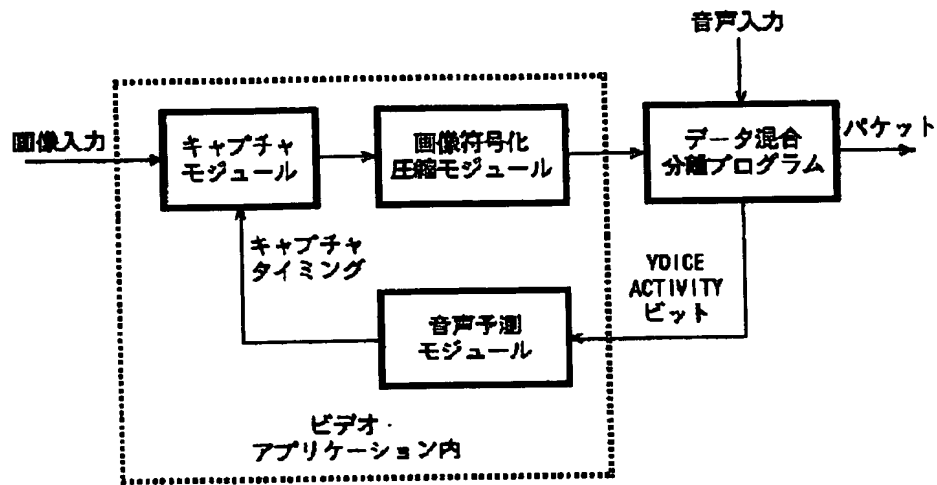


【図4】

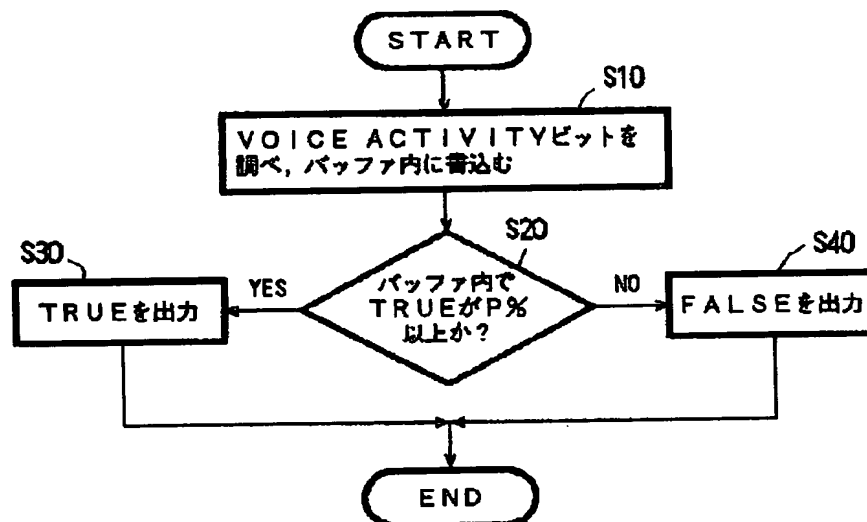




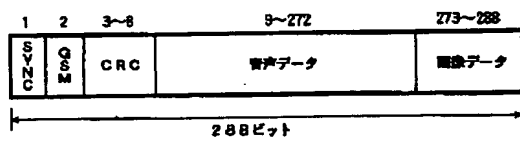
【図6】



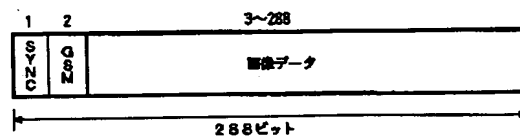
【図7】



【図10】



(a)



(b)

【図8】

